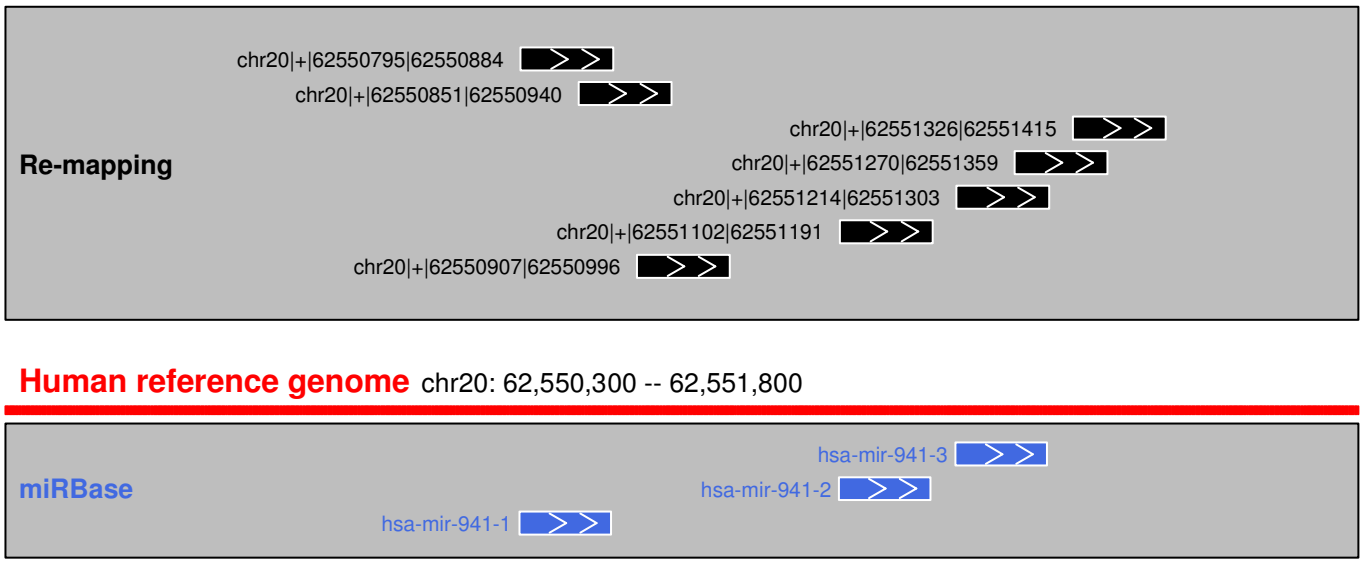
Evolution of the human-specific microRNA miR-941

Supplementary Information

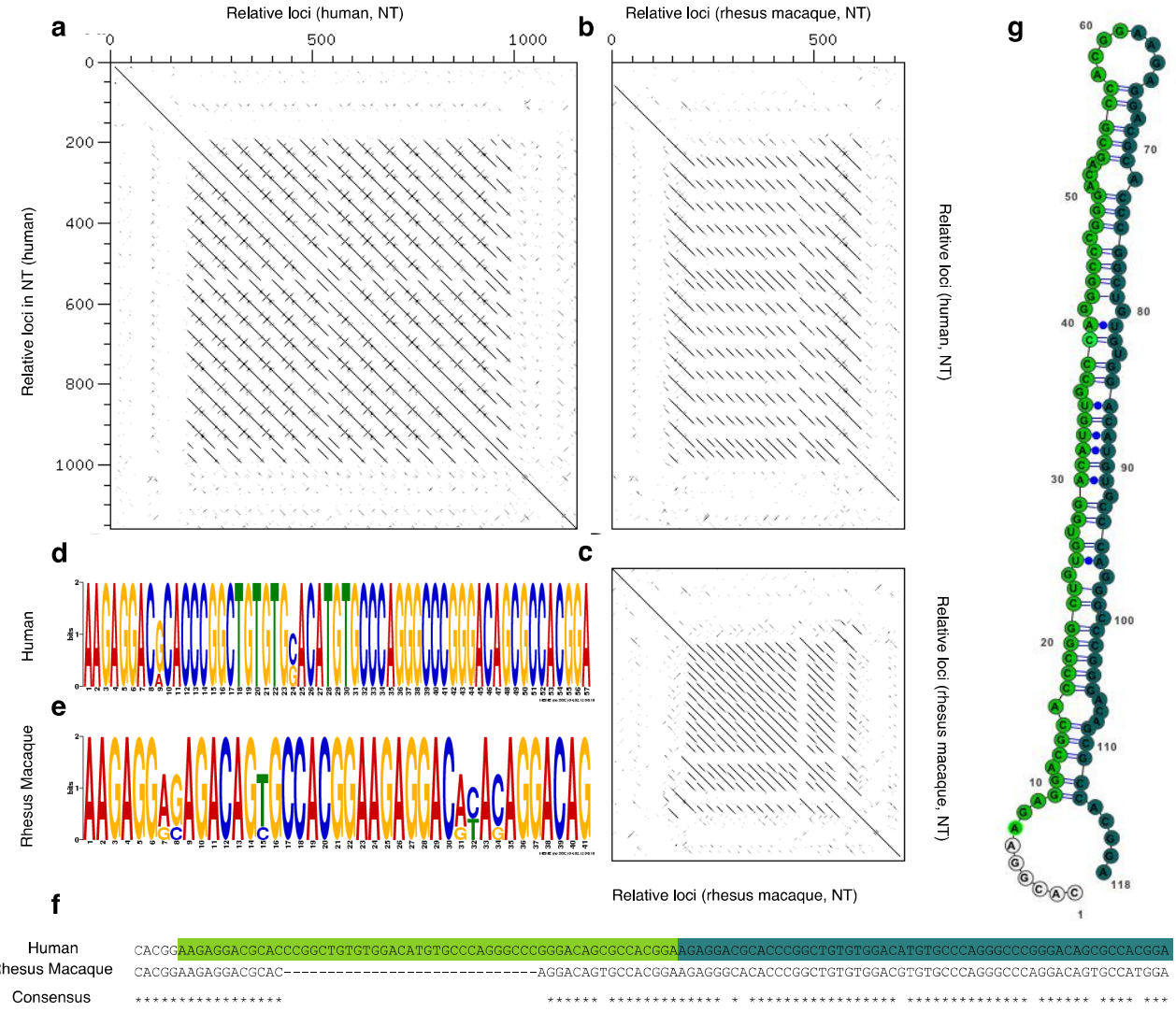
Supplementary Figures

Supplementary Figure S1: miR-941 precursor locations in the human reference genome



Re-mapping miR-941 precursor sequences to the human genome indicated that the human reference genome (UCSC genome accession code hg19) contains seven copies of the putative miR-941 precursor (black bars, upper panel). For comparison, miR-941 precursor locations (blue bars) based on miRBase annotation (version 17) are shown on the lower panel.

Supplementary Figure S2: miR-941 precursor sequence evolution



(a) Human miR-941 cluster region dot-plotted against itself. Regions of high sequence identity are shown as dark diagonal lines. The main diagonal shows self-matches, other prominent diagonals show the pattern of direct repeat matches. Prepared with Dotter (Sonnhammer and Durbin, 1995; sliding-window=18). (b) Shows human miR-941 cluster region versus macaque miR-941 cluster orthologous region dot plot. (c) Shows macaque versus macaque miR-941 cluster orthologous region dot plot. (d) and (e) Show tandem repeat sequences in human (57nt) and macaque (41nt), respectively. Tandem repeat sequences were detected by equicktandem and further illustrated with sequence logo. (f) Shows miR-941 precursor sequence alignments between human and macaque. Two human tandem repeat copies were colored in light and dark green. (g) Human miR-941 precursor secondary structure with two tandem repeats colored as in panel f.

Supplementary Figure S3: Confirmation of miR-941 genomic region sequence by PCR

(a) Human PCR product and genome sequence alignment

```
hg19_ref_dna      GGAGAGGACGCACCCGGCTGTGTGGACATGTGCCCAG-GGCCCAGGACAGCGCCACGGAA
1                 GGAGAGGACGCACCCGGCTGTGTGGACATGTGCCCAgGGGCCCAGGtCAGCGCCcCGGAA
                  ***********************************  ******** ******* *****

hg19_ref_dna      GAGGACACACCCGGCTGTGTGGACATGTGCCCAGGGCCCGGGACAGCGCCACGGAAGAGG
1                 GAGGACACACCCGGCTGTGTGGACATGTGCCCAGGGCCCGGGACAGCGCCACGGAAGAGG
                  ************************************************************

hg19_ref_dna      ACGCACCCGGCTGTGTGCACATGTGCCCAGGGCCCGGGACAGCGCCACGGAAGAGGACGC
1                 ACGCACCCGGCTGTGTGCACATGTGCCCAGGGCCCGGGACAGCGCCACGGAAGAGGACGC
                  ************************************************************

hg19_ref_dna      ACCCGGCTGTGTGCACATGTGCCCAGGGCCCGGGACAGCGCCACGGAAGAGGACGCACCC
1                 ACCCGGCTGTGTGCACATGTGCCCAGGGCCCGGGACAGCGCCACGGAAGAGGACGCACCC
                  ************************************************************

hg19_ref_dna      GGCTGTGTGCACATGTGCCCAGGGCCCGGGACAGCGCCACGGAAGAGGACGCACCCGGCT
1                 GGCTGTGTGCACATGTGCCCAGGGCCCGGGACAGCGCCACGGAAGAGGACGCACCCGGCT
                  ************************************************************

hg19_ref_dna      GTGTGGACATGTGCCCAGGGCCCGGGACAGCGCCACGGAAGAGGACGCACAGGACAGCGC
1                 GTGTGGACATGTGCCCAGGGCCCGGGACAGCGCCACGGAAGAGGACGCACAGGACAGCGC
                  ************************************************************

hg19_ref_dna      CACGGAAGAGGACGCACCCGGCTGTGTGCACATGTGCCCAGGGCCCGGGACAGCGCCACG
1                 CACGGAAGAGGACGCACCCGGCTGTGTGCACATGTGCCCAGGGCCCGGGACAGCGCCACG
                  ************************************************************

hg19_ref_dna      GAAGAGGACGCACCCGGCTGTGTGCACATGTGCCCAGGGCCCGGGACAGCGCCATGGAAG
1                 GAAGAGGACGCACCCGGCTGTGTGCACATGTGCCCAGGGCCCGGGACAGCGCCATGGAAG
                  ************************************************************

hg19_ref_dna      AGGACGCACCCGGCTGTGTGCACATGTGCCCAGGGCCCGGGACAGCGCCACGGAAGAGGA
1                 A-GACGCACCCGGCTGTGTGCACATGTGCCCAGGGCCCGGGACAGCGCCACGGAAGAGGA
                  * ***********************************************************

hg19_ref_dna      CGCACCCGGCTGTGTGCACATGTGCCCAGGGCCCGGGACAGCGCCACGGAAGAGGATGCA
1                 CGCACCCGGCTGTGTGCACATGTGCCCAGGGCCCGGGACAGCGCCACGGAAGAGGAcGCA
                  ********************************************************* ***

hg19_ref_dna      CCCGGCTGTGTGGACATGTGCCCAGGGCCCGGGACAGCGCCACGGAAGAGGACGCACCCG
1                 CCCGGCTGTGTGGACATGTGCCCAGGGCCCGGGACAGCGCCACGGAAGAGGACGCACCCG
                  ************************************************************

hg19_ref_dna      GCTGTGTGGACATGTGCCCAGGGCCCGGGACAGCGCCACGGAAGAGGACGCACCCGGCTG
1                 GCTGTGTGGACATGTGCCCAGGGCCCGGGACAGCGCCACGGAAGAGGACGCACCCGGCTG
                  ************************************************************

hg19_ref_dna      TGTGGACATGTGCCCAGGGCCCGGGACAGCGCCACGGAAGAGGACGCACAGGACAGCGCC
1                 TGTGGACATGTGCCCAGGGCCCGGGACAGCGCCACGGAAGAGGACGCACAGGACAGCGCC
                  ************************************************************
```

```
hg19_ref_dna      ACGGAAGAGGACGCACCCGGCTGTGTGGCAGGGGAGGTTCCCAAGAGCAGGCTCAGGGCT
1                 ACGGAAGAGGACGCACCCGGCTGTGTGGCAGGGGAGGTTCCCAAGAGCAGGCTCAGGGCT
                  ************************************************************

hg19_ref_dna      CTGGGGTCTGAGTGAGGGCTCTCCAGCCAGCAGGGAGCAGGGCAGGGTGTTGAAGAGACG
1                 CTGGGGTCTGAGTGAGGGCTCTCCAGCCAGCAGGGAGCAGGGCAGGGTGTTGAAGAGACG
                  ************************************************************

hg19_ref_dna      GCAGGTGCGTCAGGGGCTGACGTGGGACCCGGCACTCTGTGCTC
1                 GCAGGTGCGTCAGGGGCTGACGTGGACCCGGCACTCTGGCTC--
                  ***********************..** * ..    *  *
```

(b) Chimpanzee PCR products and genome sequence alignment (PCR result from 8 Chimpanzees)

```
panTro2_ref_dna       -----TCCCGGTCCGACGCAGGACTACCTGTTCCCGGGCAGACACAGAGCACAGAGTGCC
1                     ----TTCCCGGTCCGACGCAGGACTACCTGTTCCCGGGCAGACACAGAGCACAGAGTGCC
2                     GCTGGACCCGGTCCGACGCAGGACTACCTGTTCCCGGGCAGACACAGAGCACAGAGTGCC
3                     --------CGGTCCGACGCAGGACTACCTGTTCCCGGGCAGACACAGAGCACAGAGTGCC
4                     -----TCCCGGTCCGACGCAGGACTACCTGTTCCCGGGCAGACACAGAGCACAGAGTGCC
5                     ------TCCCGGTCCGAGCAGGACTACCTGTTCCCGGGCAGACACAGAGCACAGAGTGCC
6                     -----TCCCGGTCCGACGCAGGACTACCTGTTCCCAGGCAGACACAGAGCACAGAGTGCC
7                     -----TCCCGGTCCGACGCAGGACTACCTGTTCCCGGGCAGACACAGAGCACAGAGTGCC
8                     ----TTCCCGGTCCGACGCAGGACTACCTGTTCCCGGGCAGACACAGAGCACAGAGTGCC
                             * *  * ..*****************.**********************

panTro2_ref_dna       GGGTCCCACGTCAGCCCCTGACGCACCTGCTGTCTCTTCAACACCGTGCCCTGCTCCCTG
1                     GGGTCCCACGTCAGCCCCTGACGCACCTGCTGTCTCTTCAACACCGTGCCCTGCTCCCTG
2                     GGGTCCCACGTCAGCCCCTGACGCACCTGCTGTCTCTTCAACACCGTGCCCTGCTCCCTG
3                     GGGTCCCACGTCAGCCCCTGACGCACCTGCTGTCTCTTCAACACCGTGCCCTGCTCCCTG
4                     GGGTCCCACGTCAGCCCCTGACGCACCTGCTGTCTCTTCAACACCGTGCCCTGCTCCCTG
5                     GGGTCCCACGTCAGCCCCTGACGCACCTGCTGTCTCTTCAACACCGTGCCCTGCTCCCTG
6                     GGGTCCCACGTCAGCCCCTGACGCACCTGCTGTCTCTTCAACACTGTGCCCTGCTCCCTG
7                     GGGTCCCACGTCAGCCCCTGACGCACCTGCTGTCTCTTCAACACCGTGCCCTGCTCCCTG
8                     GGGTCCCACGTCAGCCCCTGACGCACCTGCTGTCTCTTCAACACCGTGCCCTGCTCCCTG
                      ****************************************** ***************

panTro2_ref_dna       CTGGCTGGAGAGCCCTCACTCAGACCCCAGAGCCCTGAGCCTGCCCTTGGGAACCTCTCC
1                     CTGGCTGGAGAGCCCTCACTCAGACCCCAGAGCCCTGAGCCTGCCCTTGGGAACCTCTCC
2                     CTGGCTGGAGAGCCCTCACTCAGACCCCAGAGCCCTGAGCCTGCCCTTGGGAACCTCTCC
3                     CTGGCTGGAGAGCCCTCACTCAGACCCCAGAGCCCTGAGCCTGCCCTTGGGAACCTCTCC
4                     CTGGCTGGAGAGCCCTCACTCAGACCCCAGAGCCCTGAGCCTGCCCTTGGGAACCTCTCC
5                     CTGGCTGGAGAGCCCTCACTCAGACCCCAGAGCCCTGAGCCTGCCCTTGGGAACCTCTCC
6                     CTGGCTGGAGAGCCCTCACTCAGACCCCAGAGCCCTGAGCCTGCCCTTGGGAACCTCTCC
7                     CTGGCTGGAGAGCCCTCACTCAGACCCCAGAGCCCTGAGCCTGCCCTTGGGAACCTCTCC
8                     CTGGCTGGAGAGCCCTCACTCAGACCCCAGAGCCCTGAGCCTGCCCTTGGGAACCTCTCC
                      ************************************************************

panTro2_ref_dna       TGCCACACAGCCGGGTACGTCCTCTTCCGTGGCGCTGTCCTGTGCGTCCTCTTCCGTGGC
1                     TGCCACACAGCCGGGTACGTCCTCTTCCGTGGCGCTGTCCTGTGCGTCCTCTTCCGTGGC
2                     TGCCACACAGCCGGGTACGTCCTCTTCCGTGGCGCTGTCCTGTGCGTCCTCTTCCGTGGC
3                     TGCCACACAGCCGGGTACGTCCTCTTCCGTGGCGCTGTCCTGTGCGTCCTCTTCCGTGGC
4                     TGCCACACAGCCGGGTACGTCCTCTTCCGTGGCGCTGTCCTGTGCGTCCTCTTCCGTGGC
5                     TGCCACACAGCCGGGTACGTCCTCTTCCGTGGCGCTGTCCTGTGCGTCCTCTTCCGTGGC
```

```
6                   TGCCACACAGCCGGGTACGTCCTCTTCCGTGGCGCTGTCCTGTGCGTCCTCTTCCGTGGC
7                   TGCCACACAGCCGGGTACGTCCTCTTCCGTGGCGCTGTCCTGTGTGTCCTCTTCCGTGGC
8                   TGCCACACAGCCGGGTACGTCCTCTTCCGTGGCGCTGTCCTGTGCGTCCTCTTCCGTGGC
                    ************************************** **************

panTro2_ref_dna     GCTGTCCTGGGCCCTGGGCACATGTGCACACAGCCGGGTGCATCCTCTCCCCGGACACGT
1                   GCTGTCCTGGGCCCTGGGCACATGTGCACACAGCCGGGTGCATCCTCTCCCCGGACACGT
2                   GCTGTCCTGGGCCCTGGGCACATGTGCACACAGCCGGGTGCATCCTCTCCCCGGACACGT
3                   GCTGTCCTGGGCCCTGGGCACATGTGCACACAGCCGGGTGCATCCTCTCCCCGGACACGT
4                   GCTGTCCTGGGCCCTGGGCACATGTGCACACAGCCGGGTGCATCCTCTCCCCGGACACGT
5                   GCTGTCCTGGGCCCTGGGCACATGTGCACACAGCCGGGTGCATCCTCTCCCCGGACACGT
6                   GCTGTCCTGGGCCCTGGGCACATGTGCACACAGCCGGGTGCATCCTCTCCCCGGACACGT
7                   GCTGTCCTGGGCCCTGGGCACATGTGCACACAGCCGGGTGCATCCTCTCCCCGGACACGT
8                   GCTGTCCTGGGCCCTGGGCACATGTGCACACAGCCGGGTGCATCCTCTCCCCGGACACGT
                    ************************************************************

panTro2_ref_dna     CGGTGCCATCACAGCTC---
1                   CGGTGCCATCACAGCAA---
2                   CGGTGCCATCACAGCCCACG
3                   CGGTGCCATCACAGCA----
4                   CGGTCCATCACAGCCC----
5                   CGGTCCCATCACAGC-----
6                   CGGTCCCATCACAGCT----
7                   CGGTGCCATCACAGC-----
8                   CGGTgCCCATCACAGCAA--

                    **** *.   ..
```

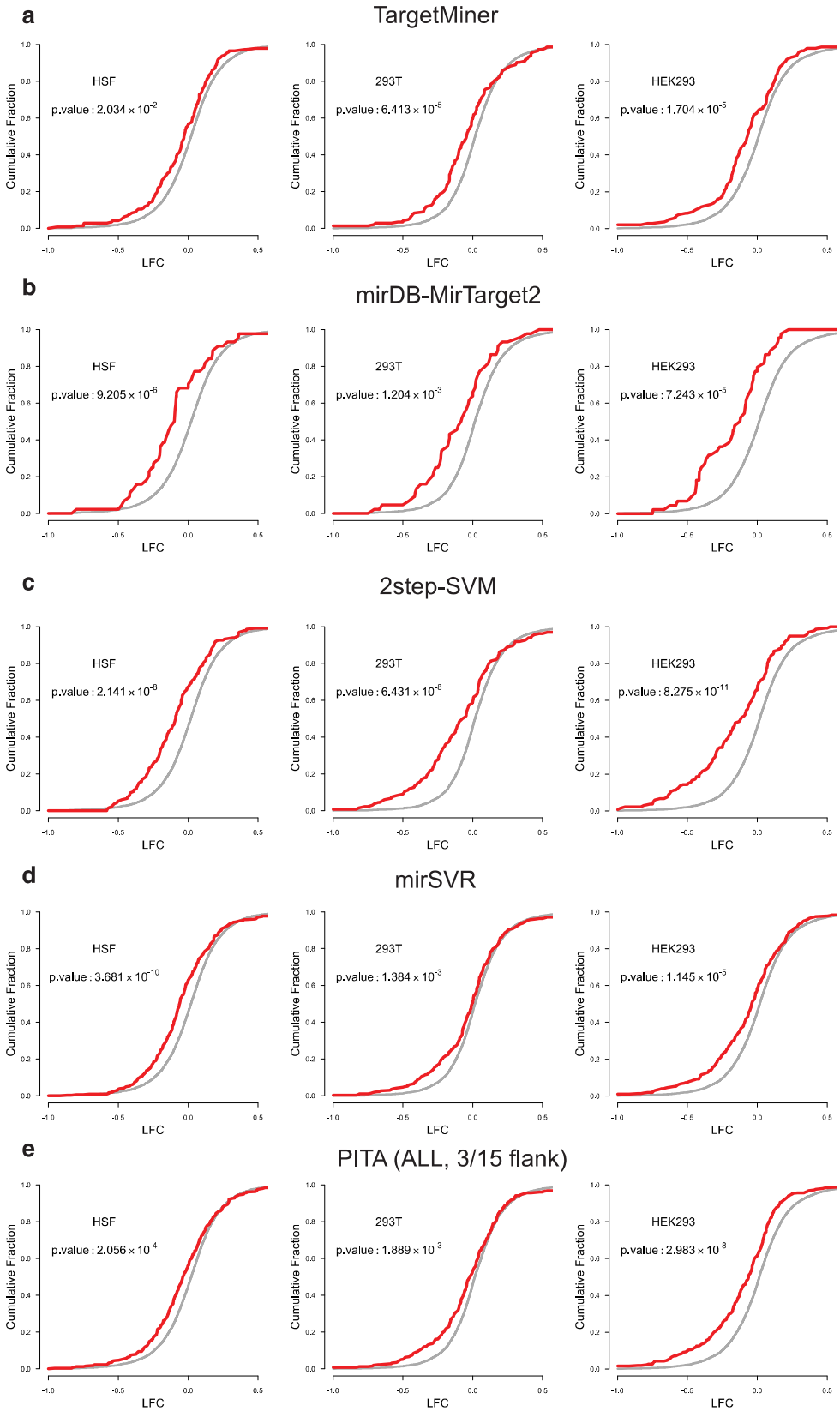
(c) Rhesus macaque PCR products and genome sequence alignment (PCR result from 6 Macaques)

```
rheMac2_ref_dna     CTGCCCTTGGGAACCTCTTCTGCCACACAGCAGGGTGCATCCTCTTCCATGGCACTGTCC
1                   CTGCCCTTGGGAACCTCTTCTGCCACACAGCAGGGTGCATCCTCTTCCATGGCACTGTCC
2                   CTGCCCTTGGGAACCTCTTCTGCCACACAGCAGGGTGCATCCTCTTCCATGGCACTGTCC
3                   CTGCCCTTGGGAACCTCTTCTGCCACACAGCAGGGTGCATCCTCTTCCATGGCACTGTCC
4                   CTGCCCTTGGGAACCTCTTCTGCCACACAGCAGGGTGCATCCTCTTCCATGGCACTGTCC
5                   CTGCCCTTGGGAACCTCTTCTGCCACACAGCAGGGTGCATCCTCTTCCATGGCACTGTCC
6                   CTGCCCTTGGGAACCTCTTCTGCCACACAGCAGGGTGCATCCTCTTCCATGGCACTGTCC
                    ************************************************************

rheMac2_ref_dna     TGGGCCCTGGGCACACGTCCACACAGCCGGGTGTGCCCTCTTCCGTGGCACTGTCCTGTG
1                   TGGGCCCTGGGCACACGTCCACACAGCCGGGTGTGCCCTCTTCCGTGGCACTGTCCTGTG
2                   TGGGCCCTGGGCACACGTCCACACAGCCGGGTGTGCCCTCTTCCGTGGCACTGTCCTGTG
3                   TGGGCCCTGGGCACACGTCCACACAGCCGGGTGTGCCCTCTTCCGTGGCACTGTCCTGTG
4                   TGGGCCCTGGGCACACGTCCACACAGCCGGGTGTGCCCTCTTCCGTGGCACTGTCCTGTG
5                   TGGGCCCTGGACACACGTCCACACAGCCGGGTGTGCCCTCTTCCGTGGCACTGTCCTGTG
6                   TGGGCCCTGGGCACACGTCCACACAGCCGGGTGTGCCCTCTTCCGTGGCACTGTCCTGTG
                    **********.************************************************

rheMac2_ref_dna     CGTCCTCTTCCGTGGCACTGTCTGTCCTCTTCTGTGGCACTATTCTGGGCCCTGGGCATA
1                   CGTCCTCTTCCGTGGCACTGTCTGTCCTCTTCTGTGGCACTATTCTGGGCCCTGGGCATA
2                   CGTCCTCTTCCGTGGCACTGTCTGTCCTCTTCTGTGGCACTATTCTGGGCCCTGGGCATA
```

```
3                   CGTCCTCTTCCGTGGCACTGTCTGTCCTCTTCTGTGGCACTATTCTGGGCCCTGGGCATA
4                   CGTCCTCTTCCGTGGCACTGTCTGTCCTCTTCTGTGGCACTATTCTGGGCCCTGGGCATA
5                   CGTCCTCTTCCGTGGCACTGTCTGTCCTCTTCTGTGGCACTATTCTGGGCCCTTGGCATA
6                   CGTCCTCTTCCGTGGCACTGTCTGTCCTCTTCTGTGGCACTATTCTGGGCCCTTGGCATA
                    ************************************************** ******

rheMac2_ref_dna     TATCAGTGTGTCCTCTTCCGTGGCGCTGTCCTGTATGTCCTCTTCCGTAGCGCTGTCCTG
1                   TATCAGTGTGTCCTCTTCCGTGGCGCTGTCCTGTATGTCCTCTTCCGTAGCGCTGTCCTG
2                   TATCAGTGTGTCCTCTTCCGTGGCGCTGTCCTGTATGTCCTCTTCCGTAGCGCTGTCCTG
3                   TATCAGTGTGTCCTCTTCCATGGCGCTGTCCTGTGTGTCCTCTTCCGTGGCACTCT----
4                   TATCAGTGTGTCCTCTTCCGTGGCGCTGTCCTGTATGTCCTCTTCCGTAGCGCTGTCCTG
5                   TATCAGTGTGTCCTCTTCCGTGGCGCTGTCCTGTATGTCCTCTTCCGTAGCGCTGTCCTG
6                   TATCAGTGTGTCCTCTTCCGTGGCGCTGTCCTGTATGTCCTCTTCCGTAGCGCTGTCCTG
                    ******************.**************.*************.**.** *

rheMac2_ref_dna     TGTGTCCTCTTCCGTGGCACTGTCCTGTGTGTCCTCTTCCGTGGCACTGTCTCTCCTCTT
1                   TGTGTCCTCTTCCGTGGCACTGTCCTGTGTGTCCTCTTCCGTGGCACTGTCTCTCCTCTT
2                   TGTGTCCTCTTCCGTGGCACTGTCCTGTGTGTCCTCTTCCGTGGCACTGTCTCTCCTCTT
3                   ---CTCCTCTTCCGTGGCGCTGTCCTGTATGTCCTCTTCCGTGGCGCTGTCTCTCCTCTT
4                   TGTGTCCTCTTCCGTGGCACTGTCCTGTGTGTCCTCTTCCTTGGCAGTGTCTCTCCTCTT
5                   TGTGTCCTCTTCCGTGGCACTGTCCTGTGTGTCCTCTTCCGTGGCACTGTCTCTCCTCTT
6                   TGTGTCCTCTTCCGTGGCACTGTCCTGTGTGTCCTCTTCCGTGGCACTGTCTCTCCTCTT
                        **************.*********.********** ****. *************

rheMac2_ref_dna     CCGTGGCGCTGTCCTGTATGTCCTCTTCCGTGGCACTGTCTCTCCTCTTCCGTGGCGCTG
1                   CCGTGGCGCTGTCCTGTATGTCCTCTTCCGTGGCGCTGTCTCTCCTCTTCCCTGGCGCTG
2                   CCGTGGCGCTGTCCTGTATGTCCTCTTCCGTGGCGCTGTCTCTCCTCTTCCGTGGCGCTG
3                   CCGTGGCGCTGTCCTGTGTGTC----------------------CTCTTCCGTGGCGCTG
4                   CCGTGGCACTGTCCTCTGTGTCCTCTTCCTTGGCACTGTCTCTCCTCTTCCGTGGCTCTG
5                   CCGTGGCGCTGTCCTGTATGTCCTCTTCCGTGGCACTGTCTCTCCTCTTCCGTGGCGCTG
6                   CCGTGGCGCTGTCCTGTATGTCCTCTTCCGTGGCGCTGTCTCTCCTCTTCCGTGGCGCTG
                    *******.******* *.****                     ******* **** ***

rheMac2_ref_dna     TCCTG---------------------------TATGTCCTCTTCCGTGGCGCTGTCTCCC
1                   TCCTGTGTGTCCTCTTCCGTGGCACTGTCCTGTGTGTCCTCTTCCGTGGCACTGTCTCTC
2                   TCCTGTGTGTCCTCTTCCGTGGCACTGTCCTGTGTGTCCTCTTCCGTGGCACTGTCTCTC
3                   TCCTGTGTGTCCTCTTCCGTGGCGCTGTCCTGTGTGTCCTCTTCCGTGGCACTGTCTCTC
4                   TCCTG---------------------------TATGTCCTCTTCCGTGGCGCTGTCTCGC
5                   TCCTGTGTGTCCTCTTCCGTGGCACTGTCCTGTGTGTCCTCTTCCGTGGCACTGTCTCTC
6                   TCCTGTGTGTCCTCTTCCGTGGCGCTGTCCTGTGTGTCCTCTTCCGTGGCACTGTCTCTC
                    *****                          *.****************.******* *

rheMac2_ref_dna     CTCTTCCGTGGCACTGTCCTCTGTGTCCTCTTCCGTGGCACTGTCTCTCCTCTTCCGTGG
1                   CTCTTCCGTGGCACTGTCCTCTGTGTCCTCTTCCGTGGCACTGTCTCTCCTCTTCCGTGG
2                   CTCTTCCGTGGCACTGTCCTCTGTGTCCTCTTCCGTGGCACTGTCTCTCCTCTTCCGTGG
3                   CTCTTCCGTGGCACTGTCCTCTGTGTCCTCTTCCGTGGCACTGTCTCTCCTCTTCCGTGG
4                   CTCTTCCGTGGCACTGTCCTCTGTGTCCTCTTCCGTGGCACTGTCTCGCCTCTTCTGTGG
5                   CTCTTCCGTGGCACTGTCCTCTGTGTCCTCTTCCGTGGCACTGTCTCTCCTCTTCCGTGG
6                   CTCTTCCGTGGCACTGTCCTCTGTGTCCTCTTCCGTGGCACTGTCTCTCCTCTTCCGTGG
                    ********************************************** ******* ****

rheMac2_ref_dna     CGCTGTCCTGGGCGCTGGGCACACATCCACACGGCAGGGTGCATCCTCTCCCTGGACACA
1                   CGCTGTACTGGGCGCTGGGCCAACATCCACACAGCAGGGTGCATCCTCTCCCTGGACACA
2                   CGCTGTACTGGGCGCTGGGCCAACATCCACACGGCAGGGTGCATCCTCTCCCTGGACACA
3                   CGCTGTCCTGGGCGCTGGGCACACATCCACACGGCAGGGTGCATCCTCTCCCTGGACACA
4                   CGCTGTCCTGGGCGCTGGGCACACATCCACACGGCAGGGTGCATCCTCTCCCTGGACACA
5                   CGCTGTCCTGGGCGCTGGGCACACATCCACACGGCAGGGTGCATCCTCTCCCTGGACAAA
6                   CGCTGTCCTGGGCGCTGGGCACACATCCACACGGCAGGGTGCATCCTCTCCCTGGACACA
                    ******.*************..**********.************************.*
```

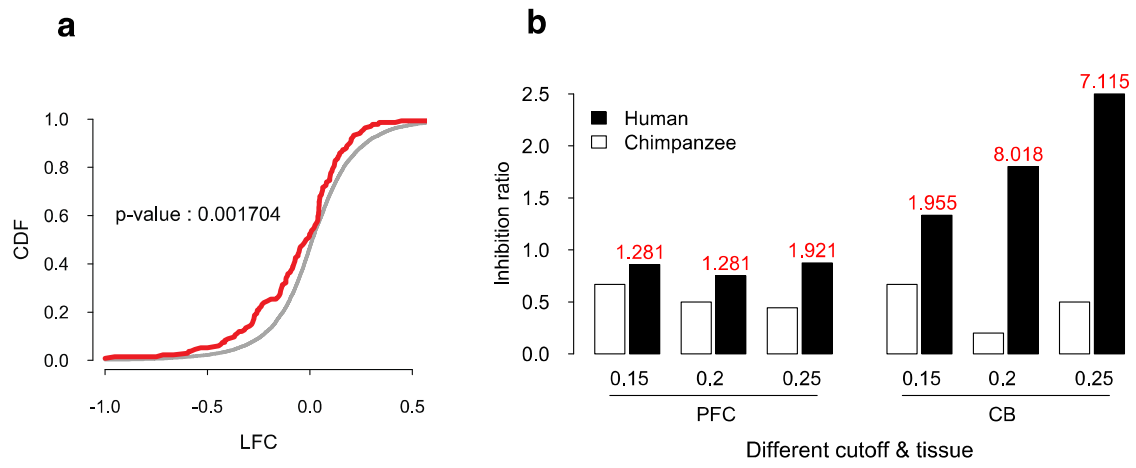
Sequence alignments between miR-941 genomic region in the reference genomes of human (UCSC genome accession code hg19) (a), chimpanzee (UCSC genome accession code panTro2) (b), and rhesus macaque (UCSC genome accession code rheMac2) (c) and sequences of the corresponding genomic regions PCR amplified from these species' DNA samples (one human, eight chimpanzees and six macaques). The alignments were built using Clustal Omega tool. Human miR-941 cluster region: chr20: 62550779-62551468 (690nt, UCSC genome accession code hg19), human PCR target region: chr20: 62550726-62,5516 68 (943nt, UCSC genome accession code hg19), Chimpanzee PCR target region: chr20: 61884228-61884539 (312nt, UCSC genome accession code panTro2) corresponding to the human genome region chr20: 62550699-62551709 (1011nt, UCSC genome accession code hg19), Macaque PCR target region: chr10: 424154-424666 (513nt, UCSC genome accession code rheMac2) corresponding to the human genome region chr20: 62550718-62551554 (837nt, UCSC genome accession code hg19).

Supplementary Figure S4: miRNA-941 target effects in the human cell lines



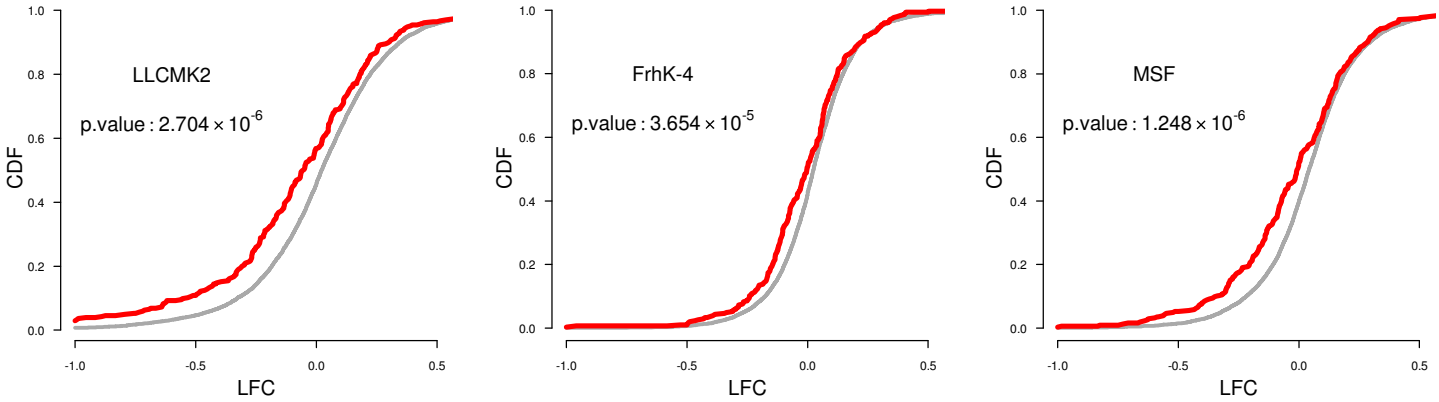
Cumulative distribution plots of log2-transformed gene expression fold-changes (LFC) for genes containing miR-941 target sites (red) predicted by TargetMiner (a), mirDB-MirTarget2 (b), 2step-SVM (c), mirSVR (d) and PITA (ALL, 3/15 flank) (e) and all the other expressed genes (grey) in three human cells lines: 293T, HEK293 and HSF2, transfected with miR-941 duplex or mock duplex. The y-axis shows cumulative distribution function (CDF) of LFC distribution. The *p*-values were calculated by Kolmogorov-Smirnov test.

Supplementary Figure S5: miRNA-941 targets identification using Ago2-IP method



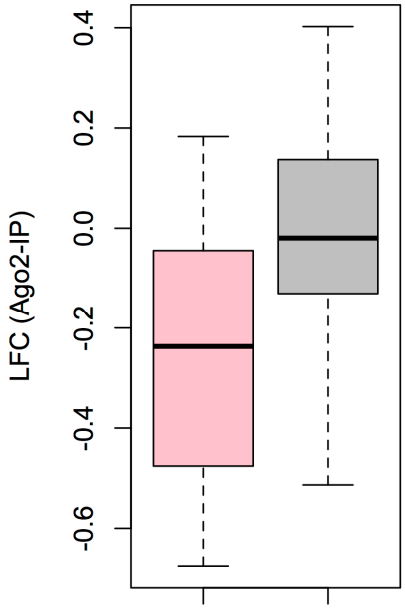
(a) Cumulative distribution plots of log2-transformed gene expression fold-changes (LFC) for target genes obtained using Ago2-IP experiment in 293T cell line (red) and all other expressed genes (grey) after transfection with miR-941 duplex or mock duplex in the human 293T cells line. Prevalence of the negative LFC measurements among miR-941 targets obtained using Ago2-IP experiment indicates targets obtained by miR-941 transfection experiment were direct targets. The y-axis shows cumulative distribution function (CDF) of LFC distribution. (b) Transcriptional inhibition of experimentally verified miR-941 target genes using Ago2-IP experiment in the prefrontal cortex (PFC) and cerebellum (CB). The x-axis shows log2-transformed gene expression fold-changes (LFC) cutoffs, LFC were calculated between Ago2-IP and control experiments. The numbers in red were odds ratio of transcriptional inhibition between human and chimpanzee at each LFC cutoffs.

Supplementary Figure S6: miRNA-941 target effects in rhesus macaque cell lines



Cumulative distribution plots of log2-transformed gene expression fold-changes (LFC) for genes containing miR-941 target sites predicted by TargetScan (red) and all the others (grey) in three rhesus macaque cells lines: LLCMK2, FrhK-4 and MSF, transfected with miR-941 duplex or mock duplex. The y-axis shows cumulative distribution function (CDF) of LFC distribution. The *p*-values were calculated by Kolmogorov-Smirnov test. Predictions by other tools were not applicable to the macaque genome.

Supplementary Figure S7: The Ago2-IP enrichment of miR-941 target genes showing miR-941 inhibition avoidance in human, but not in macaque cell lines



Shown are the distributions of $\log_2$ fold-change (LFC) values in Ago2-IP experiments for 19 human genes showing miR-941 inhibition avoidance in human, but not in macaque cell lines (pink) and all 124 miR-941 predicted target genes expressed in human and macaque cells and containing miR-941 binding sites in both species. The LFC values were determined based on differences in transcript abundance observed in Ago2-IP experiment conducted in human 293T cell line between cells transfected with miR-941 duplex and cells transfected with a mock duplex. Negative LFC values correspond to target depletion in Ago2-IP complexes in cells transfected with miR-941 duplex compared to the cells transfected with mock. Genes showing inhibition avoidance were defined as genes containing miR-941 binding sites in humans and macaques and showing expression down-regulation after miR-941 duplex transfection in macaque cell lines, but escaping down-regulation in human cell lines after transfection with miR-941 duplex. The 19 genes showing such miR-941 inhibition avoidance are significantly under-represented in Ago2-IP complexes obtained after transfection of miR-941 duplex in 293T cells compared to other predicted miR-941 target genes (Wilcoxon signed-rank test, $p<0.001$).

Supplementary Figure S8: Sequence alignments of miR-941 genomic region PCR product between independent PCR replicates from 6 human individuals

```
individual3               CACCCGGCTGTGTGCACATGTGCCCAGGGCCCGGGACAGCGCCACGGAAGAGGACGCACC
individual3_replicate     CACCCGGCTGTGTGCACATGTGCCCAGGGCCCGGGACAGCGCCACGGAAGAGGACGCACC
individual4               CACCCGGCTGTGTGCACATGTGCCCAGGGCCCGGGACAGCGCCACGGAAGAGGACGCACC
individual4_replicate     CACCCGGCTGTGTGCACATGTGCCCAGGGCCCGGGACAGCGCCACGGAAGAGGACGCACC
individual5               CACCCGGCTGTGTGCACATGTGCCCAGGGCCCGGGACAGCGCCACGGAAGAGGACGCACC
individual5_replicate     CACCCGGCTGTGTGCACATGTGCCCAGGGCCCGGGACAGCGCCACGGAAGAGGACGCACC
individual6               CACCCGGCTGTGTGCACATGTGCCCAGGGCCCGGGACAGCGCCACGGAAGAGGACGCACC
individual6_replicate     CACCCGGCTGTGTGCACATGTGCCCAGGGCCCGGGACAGCGCCACGGAAGAGGACGCACC
                          ************************************************************
                          ----------------------
individual1               CGGCTGTGTGGACATGTGCCCAGGGCCCGGGACAGCGCCACGGAAGAGGACGCACAGGAC
individual1_replicate     CGGCTGTGTGGACATGTGCCCAGGGCCCGGGACAGCGCCACGGAAGAGGACGCACAGGAC
individual2               CGGCTGTGTGGACATGTGCCCAGGGCCCGGGACAGCGCCACGGAAGAGGACGCACAGGAC
individual2_replicate     CGGCTGTGTGGACATGTGCCCAGGGCCCGGGACAGCGCCACGGAAGAGGACGCACAGGAC
individual3               CGGCTGTGTGGACATGTGCCCAGGGCCCGGGACAGCGCCACGGAAGAGGACGCACAGGAC
individual3_replicate     CGGCTGTGTGGACATGTGCCCAGGGCCCGGGACAGCGCCACGGAAGAGGACGCACAGGAC
individual4               CGGCTGTGTGGACATGTGCCCAGGGCCCGGGACAGCGCCACGGAAGAGGACGCACAGGAC
individual4_replicate     CGGCTGTGTGGACATGTGCCCAGGGCCCGGGACAGCGCCACGGAAGAGGACGCACAGGAC
individual5               CGGCTGTGTGGACATGTGCCCAGGGCCCGGGACAGCGCCACGGAAGAGGACGCACAGGAC
individual5_replicate     CGGCTGTGTGGACATGTGCCCAGGGCCCGGGACAGCGCCACGGAAGAGGACGCACAGGAC
individual6               CGGCTGTGTGGACATGTGCCCAGGGCCCGGGACAGCGCCACGGAAGAGGACGCACAGGAC
individual6_replicate     CGGCTGTGTGGACATGTGCCCAGGGCCCGGGACAGCGCCACGGAAGAGGACGCACAGGAC
                          ************************************************************

individual1               AGCGCCACGGAAGAGGACGCACCCGGCTGTGTGCACATGTGCCCAGGGCCCGGGACAGCG
individual1_replicate     AGCGCCACGGAAGAGGACGCACCCGGCTGTGTGCACATGTGCCCAGGGCCCGGGACAGCG
individual2               AGCGCCACGGAAGAGGACGCACCCGGCTGTGTGCACATGTGCCCAGGGCCCGGGACAGCG
individual2_replicate     AGCGCCACGGAAGAGGACGCACCCGGCTGTGTGCACATGTGCCCAGGGCCCGGGACAGCG
individual3               AGCGCCACGGAAGAGGACGCACCCGGCTGTGTGCACATGTGCCCAGGGCCCGGGACAGCG
individual3_replicate     AGCGCCACGGAAGAGGACGCACCCGGCTGTGTGCACATGTGCCCAGGGCCCGGGACAGCG
individual4               AGCGCCACGGAAGAGGACGCACCCGGCTGTGTGCACATGTGCCCAGGGCCCGGGACAGCG
individual4_replicate     AGCGCCACGGAAGAGGACGCACCCGGCTGTGTGCACATGTGCCCAGGGCCCGGGACAGCG
individual5               AGCGCCACGGAAGAGGACGCACCCGGCTGTGTGCACATGTGCCCAGGGCCCGGGACAGCG
individual5_replicate     AGCGCCACGGAAGAGGACGCACCCGGCTGTGTGCACATGTGCCCAGGGCCCGGGACAGCG
individual6               AGCGCCACGGAAGAGGACGCACCCGGCTGTGTGCACATGTGCCCAGGGCCCGGGACAGCG
individual6_replicate     AGCGCCACGGAAGAGGACGCACCCGGCTGTGTGCACATGTGCCCAGGGCCCGGGACAGCG
                          ************************************************************

individual1               CCACGGAAGAGGACGCACCCGGCTGTGTGCACATGTGCCCAGGGCCCGGGACAGCGCCAT
individual1_replicate     CCACGGAAGAGGACGCACCCGGCTGTGTGCACATGTGCCCAGGGCCCGGGACAGCGCCAT
individual2               CCACGGAAGAGGACGCACCCGGCTGTGTGCACATGTGCCCAGGGCCCGGGACAGCGCCAT
individual2_replicate     CCACGGAAGAGGACGCACCCGGCTGTGTGCACATGTGCCCAGGGCCCGGGACAGCGCCAT
individual3               CCACGGAAGAGGACGCACCCGGCTGTGTGCACATGTGCCCAGGGCCCGGGACAGCGCCAC
individual3_replicate     CCACGGAAGAGGACGCACCCGGCTGTGTGCACATGTGCCCAGGGCCCGGGACAGCGCCAC
individual4               CCACGGAAGAGGACGCACCCGGCTGTGTGCACATGTGCCCAGGGCCCGGGACAGCGCCAC
individual4_replicate     CCACGGAAGAGGACGCACCCGGCTGTGTGCACATGTGCCCAGGGCCCGGGACAGCGCCAC
individual5               CCACGGAAGAGGACGCACCCGGCTGTGTGCACATGTGCCCAGGGCCCGGGACAGCGCCAT
individual5_replicate     CCACGGAAGAGGACGCACCCGGCTGTGTGCACATGTGCCCAGGGCCCGGGACAGCGCCAT
individual6               CCACGGAAGAGGACGCACCCGGCTGTGTGCACATGTGCCCAGGGCCCGGGACAGCGCCAT
individual6_replicate     CCACGGAAGAGGACGCACCCGGCTGTGTGCACATGTGCCCAGGGCCCGGGACAGCGCCAT
                          ************************************************************
                                    ----------------------
individual1               GGAAGAGGACGCACCCGGCTGTGTGCACATGTGCCCAGGGCCCGGGACAGCGCCACGGAA
individual1_replicate     GGAAGAGGACGCACCCGGCTGTGTGCACATGTGCCCAGGGCCCGGGACAGCGCCACGGAA
individual2               GGAAGAGGACGCACCCGGCTGTGTGCACATGTGCCCAGGGCCCGGGACAGCGCCACGGAA
individual2_replicate     GGAAGAGGACGCACCCGGCTGTGTGCACATGTGCCCAGGGCCCGGGACAGCGCCACGGAA
individual3               GGAAGAGGACGCACCCGGCTGTGTGCACATGTGCCCAGGGCCCGGGACAGCGCCACGGAA
individual3_replicate     GGAAGAGGACGCACCCGGCTGTGTGCACATGTGCCCAGGGCCCGGGACAGCGCCACGGAA
individual4               GGAAGAGGACGCACCCGGCTGTGTGCACATGTGCCCAGGGCCCGGGACAGCGCCACGGAA
individual4_replicate     GGAAGAGGACGCACCCGGCTGTGTGCACATGTGCCCAGGGCCCGGGACAGCGCCACGGAA
```

```
individual5               GGAAGAGGACGCACCCGGCTGTGTGCACATGTGCCCAGGGCCCGGGACAGCGCCACGGAA
individual5_replicate     GGAAGAGGACGCACCCGGCTGTGTGCACATGTGCCCAGGGCCCGGGACAGCGCCACGGAA
individual6               GGAAGAGGACGCACCCGGCTGTGTGCACATGTGCCCAGGGCCCGGGACAGCGCCACGGAA
individual6_replicate     GGAAGAGGACGCACCCGGCTGTGTGCACATGTGCCCAGGGCCCGGGACAGCGCCACGGAA
                          ************************************************************

individual1               GAGGACGCACCCGGCTGTGTGCACATGTGCCCAGGGCCCGGGACAGCGCCACGGAAGAGG
individual1_replicate     GAGGACGCACCCGGCTGTGTGCACATGTGCCCAGGGCCTGGGACAGCGCCACGGAAGAGG
individual2               GAGGACGCACCCGGCTGTGTGCACATGTGCCCAGGGCCCGGGACAGCGCCACGGAAGAGG
individual2_replicate     GAGGACGCACCCGGCTGTGTGCACATGTGCCCAGGGCCCGGGACAGCGCCACGGAAGAGG
individual3               GAGGACGCACCCGGCTGTGTGCACATGTGCCCAGGGCCCGGGACAGCGCCACGGAAGAGG
individual3_replicate     GAGGACGCACCCGGCTGTGTGCACATGTGCCCAGGGCCCGGGACAGCGCCACGGAAGAGG
individual4               GAGGACGCACCCGGCTGTGTGCACATGTGCCCAGGGCCCGGGACAGCGCCACGGAAGAGG
individual4_replicate     GAGGACGCACCCGGCTGTGTGCACATGTGCCCAGGGCCCGGGACAGCGCCACGGAAGAGG
individual5               GAGGACGCACCCGGCTGTGTGCACATGTGCCCAGGGCCCGGGACAGCGCCACGGAAGAGG
individual5_replicate     GAGGACGCACCCGGCTGTGTGCACATGTGCCCAGGGCCCGGGACAGCGCCACGGAAGAGG
individual6               GAGGACGCACCCGGCTGTGTGCACATGTGCCCAGGGCCCGGGACAGCGCCACGGAAGAGG
individual6_replicate     GAGGACGCACCCGGCTGTGTGCACATGTGCCCAGGGCCTGGGACAGCGCCACGGAAGAGG
                          ************************************ ********************
                                ----------------------

individual1               ATGCACCCGGCTGTGTGGACATGTGCCCAGGGCCCGGGACAGCGCCACGGAAGAGGACGC
individual1_replicate     ATGCACCCGGCTGTGTGGACATGTGCCCAGGGCCCGGGACAGCGCCACGGAAGAGGACGC
individual2               ATGCACCCGGCTGTGTGGACATGTGCCCAGGGCCCGGGACAGCGCCACGGAAGAGGACGC
individual2_replicate     ATGCACCCGGCTGTGTGGACATGTGCCCAGGGCCCGGGACAGCGCCACGGAAGAGGACGC
individual3               ATGCACCCGGCTGTGTGGACATGTGCCCAGGGCCCGGGACAGCGCCACGGAAGAGGACGC
individual3_replicate     ATGCACCCGGCTGTGTGGACATGTGCCCAGGGCCCGGGACAGCGCCACGGAAGAGGACGC
individual4               ATGCACCCGGCTGTGTGGACATGTGCCCAGGGCCCGGGACAGCGCCACGGAAGAGGACGC
individual4_replicate     ATGCACCCGGCTGTGTGGACATGTGCCCAGGGCCCGGGACAGCGCCACGGAAGAGGACGC
individual5               ATGCACCCGGCTGTGTGGACATGTGCCCAGGGCCCGGGACAGCGCCACGGAAGAGGACGC
individual5_replicate     ATGCACCCGGCTGTGTGGACATGTGCCCAGGGCCCGGGACAGCGCCACGGAAGAGGACGC
individual6               ATGCACCCGGCTGTGTGGACATGTGCCCAGGGCCCGGGACAGCGCCACGGAAGAGGACGC
individual6_replicate     ATGCACCCGGCTGTGTGGACATGTGCCCAGGGCCCGGGACAGCGCCACGGAAGAGGACGC
                          ************************************************************
                             ----------------------                                  -
individual1               ACCCGGCTGTGTGGACATGTGCCCAGGGCCCGGGACAGCGCCACGGAAGAGGACGCACCC
individual1_replicate     ACCCGGCTGTGTGGACATGTGCCCAGGGCCCGGGACAGCGCCACGGAAGAGGACGCACCC
individual2               ACCCGGCTGTGTGGACATGTGCCCAGGGCCCGGGACAGCGCCACGGAAGAGGACGCACCC
individual2_replicate     ACCCGGCTGTGTGGACATGTGCCCAGGGCCCGGGACAGCGCCACGGAAGAGGACGCACCC
individual3               ACCCGGCTGTGTGGACATGTGCCCAGGGCCCGGGACAGCGCCACGGAAGAGGACGCACCC
individual3_replicate     ACCCGGCTGTGTGGACATGTGCCCAGGGCCCGGGACAGCGCCACGGAAGAGGACGCACCC
individual4               ACCCGGCTGTGTGGACATGTGCCCAGGGCCCGGGACAGCGCCACGGAAGAGGACGCACCC
individual4_replicate     ACCCGGCTGTGTGGACATGTGCCCAGGGCCCGGGACAGCGCCACGGAAGAGGACGCACCC
individual5               ACCCGGCTGTGTGGACATGTGCCCAGGGCCCGGGACAGCGCCACGGAAGAGGACGCACCC
individual5_replicate     ACCCGGCTGTGTGGACATGTGCCCAGGGCCCGGGACAGCGCCACGGAAGAGGACGCACCC
individual6               ACCCGGCTGTGTGGACATGTGCCCAGGGCCCGGGACAGCGCCACGGAAGAGGACGCACCC
individual6_replicate     ACCCGGCTGTGTGGACATGTGCCCAGGGCCCGGGACAGCGCCACGGAAGAGGACGCACCC
                          ************************************************************
                          ----------------------

individual1               GGCTGTGTGGACATGTGCCCAGGGCCCGGGACAGCGCCACGGAAGAGGACGCACAGGACA
individual1_replicate     GGCTGTGTGGACATGTGCCCAGGGCCCGGGACAGCGCCACGGAAGAGGACGCACAGGACA
individual2               GGCTGTGTGGACATGTGCCCAGGGCCCGGGACAGCGCCACGGAAGAGGACGCACAGGACA
individual2_replicate     GGCTGTGTGGACATGTGCCCAGGGCCCGGGACAGCGCCACGGAAGAGGACGCACAGGACA
individual3               GGCTGTGTGGACATGTGCCCAGGGCCCGGGACAGCGCCACGGAAGAGGACGCACAGGACA
individual3_replicate     GGCTGTGTGGACATGTGCCCAGGGCCCGGGACAGCGCCACGGAAGAGGACGCACAGGACA
individual4               GGCTGTGTGGACATGTGCCCAGGGCCCGGGACAGCGCCACGGAAGAGGACGCACAGGACA
individual4_replicate     GGCTGTGTGGACATGTGCCCAGGGCCCGGGACAGCGCCACGGAAGAGGACGCACAGGACA
individual5               GGCTGTGTGGACATGTGCCCAGGGCCCGGGACAGCGCCACGGAAGAGGACGCACAGGACA
individual5_replicate     GGCTGTGTGGACATGTGCCCAGGGCCCGGGACAGCGCCACGGAAGAGGACGCACAGGACA
individual6               GGCTGTGTGGACATGTGCCCAGGGCCCGGGACAGCGCCACGGAAGAGGACGCACAGGACA
individual6_replicate     GGCTGTGTGGACATGTGCCCAGGGCCCGGGACAGCGCCACGGAAGAGGACGCACAGGACA
```

```
                          ********************************************************

individual1               GCGCCACGGAAGAGGACGCACCCGGCTGTGTGGCAGGGGAGGTTCCCAAGAGcgggCTCA
individual1_replicate     GCGCCACGGAAGAGGACGCACCCGGCTGTGTGGCAGGGGAGGTTCCCAAGAGCAGGCTCA
individual2               GCGCCACGGAAGAGGACGCACCCGGCTGTGTGGCAGGGGAGGTTCCCAAGAGCAGGCTCA
individual2_replicate     GCGCCACGGAAGAGGACGCACCCGGCTGTGTGGCAGGGGAGGTTCCCAAGAGCAGGCTCA
individual3               GCGCCACGGAAGAGGACGCACCCGGCTGTGTGGCAGGGGAGGTTCCCAAGAGCAGGCTCA
individual3_replicate     GCGCCACGGAAGAGGACGCACCCGGCTGTGTGGCAGGGGAGGTTCCCAAGAGCAGGCTCA
individual4               GCGCCACGGAAGAGGACGCACCCGGCTGTGTGGCAGGGGAGGTTCCCAAGAGCAGGCTCA
individual4_replicate     GCGCCACGGAAGAGGACGCACCCGGCTGTGTGGCAGGGGAGGTTCCCAAGAGCAGGCTCA
individual5               GCGCCACGGAAGAGGACGCACCCGGCTGTGTGGCAGGGGAGGTTCCCAAGAGCAGGCTCA
individual5_replicate     GCGCCACGGAAGAGGACGCACCCGGCTGTGTGGCAGGGGAGGTTCCCAAGAGCAGGCTCA
individual6               GCGCCACGGAAGAGGACGCACCCGGCTGTGTGGCAGGGGAGGTTCCCAAGAGcgggCTCA
individual6_replicate     GCGCCACGGAAGAGGACGCACCCGGCTGTGTGGCAGGGGAGGTTCCCAAGAGCAGGCTCA
                          **************************************************    ****

individual1               GGGCTCTGGGGTCTGAGTGAGGGCTCTCCAGCCAGCAGggagcagggcagggtgttgaag
individual1_replicate     GGGCTCTGGGGTCTGAGTGAGGGCTCTCCAGCCAGCAGGgagcagggcagggtgttgaag
individual2               GGGCTCTGGGGTCTGAGTGAGGGCTCTCCAGCCAGCagggagcagggcagggtgttgaag
individual2_replicate     GGGCTCTGGGGTCTGAGTGAGGGCTCTCCAGCCAGCAgggagcagggcagggtgttgaag
individual3               GGGCTCTGGGGTCTGAGTGAGGGCTCTCCAGCCAGCAgggagcagggcagggtgttgaag
individual3_replicate     GGGCTCTGGGGTCTGAGTGAGGGCTCTCCAGCCAGCAgggagcagggcagggtgttgaag
individual4               GGGCTCTGGGGTCTGAGTGAGGGCTCTCCAGCCAGCAgggagcagggcagggtgttgaag
individual4_replicate     GGGCTCTGGGGTCTGAGTGAGGGCTCTCCAGCCAGCAgggagcagggcagggtgttgaag
individual5               GGGCTCTGGGGTCTGAGTGAGGGCTCTCCAGCCAGCagggagcagggcagggtgttgaag
individual5_replicate     GGGCTCTGGGGTCTGAGTGAGGGCTCTCCAGCCAGCAgggagcagggcagggtgttgaag
individual6               GGGCTCTGGGGTCTGAGTGAGGGCTCTCCAGCCAGCAGggagcagggcagggtgttgaag
individual6_replicate     GGGCTCTGGGGTCTGAGTGAGGGCTCTCCAGCCAGCAGGgagcagggcagggtgttgaag
                          ***********************************   ********************

individual1               agacggcaggtgcgtcaggggctgacgtggacccggcactcttgctct-
individual1_replicate     agacggcaggtgcgtcaggggctgacgtggacccggcactcttgctctg
individual2               agacggcaggtgcgtcaggggctgacgtggacccggcactctggct---
individual2_replicate     agacggcaggtgcgtcaggggctgacgtggacccggcactcgtgctc--
individual3               agacggcaggtgcgtcaggggctgacgtggacccggcactctggctg--
individual3_replicate     agacggcaggtgcgtcaggggctgacgtggacccggcactctggctg--
individual4               agacggcaggtgcgtcaggggctgacgtggacccggcactctggctg--
individual4_replicate     agacggcaggtgcgtcaggggctgacgtggacccggcactctggctg--
individual5               agacggcaggtgcgtcaggggctgacgtggacccggcactctggct---
individual5_replicate     agacggcaggtgcgtcaggggctgacgtggacccggcactcgtgctc--
individual6               agacggcaggtgcgtcaggggctgacgtggacccggcactcttgctct-
individual6_replicate     agacggcaggtgcgtcaggggctgacgtggacccggcactcttgctctg
                          ************************************  ***
```

Sequence alignments of miR-941 genomic region PCR product between independent PCR replicates from six human individuals. Genomic DNA from six individuals from African populations used in our miR-941 precursor region polymorphism analysis were randomly selected for replicate PCR amplification followed by sequencing using the same procedure. In all six cases, miR-941 precursor copy number estimates agreed between the experimental replicates. The genome of each individual contains 7 miR-941 mature sequences as labeled in yellow.

Supplementary Tables

Supplementary Table S1. List of human specific miRNAs

miRNA with human-specific seed region sequence changes					
miRNA id	**Precursor id**	**PFC1**	**PFC2**	**CB1**	**CB2**
hsa-miR-1229	hsa-mir-1229	0 *	0	1	0
hsa-miR-466	hsa-mir-466	0	0	0	0
hsa-miR-4804-5p	hsa-mir-4804	0	0	1	0
hsa-miR-1537	hsa-mir-1537	1	0	2	0
hsa-miR-4649-5p	hsa-mir-4649	0	1	0	0
hsa-miR-4465	hsa-mir-4465	0	0	0	0
hsa-miR-4530	hsa-mir-4530	0	0	0	0
hsa-miR-3136-3p	hsa-mir-3136	0	0	0	0
hsa-miR-3679-3p	hsa-mir-3679	0	0	0	0
hsa-miR-4322	hsa-mir-4322	0	0	0	0

Human-specific miRNA without ortholog in other 11 species					
miRNA id	**Precursor id**	**PFC1**	**PFC2**	**CB1**	**CB2**
hsa-miR-5095	hsa-mir-5095	0	0	0	0
hsa-miR-4487	hsa-mir-4487	1	0	0	0
hsa-miR-3690	hsa-mir-3690	3	4	8	11
hsa-miR-4739	hsa-mir-4739	0	0	0	0
	hsa-mir-941-1				
	has-mir-941-2				
	hsa-mir-941-3				
hsa-miR-941	hsa-mir-941-4	95	116	42	20
hsa-miR-3156-5p	hsa-mir-3156-3	0	1	0	0
hsa-miR-572	hsa-mir-572	0	0	0	0
	hsa-mir-1302-2				
	hsa-mir-1302-10				
hsa-miR-1302	hsa-mir-1302-11	4	1	5	0
hsa-miR-3673	hsa-mir-3673	0	0	0	0
hsa-miR-3648	hsa-mir-3648	0	0	0	0
hsa-miR-3911	hsa-mir-3911	0	0	1	1
hsa-miR-4781-3p	hsa-mir-4781	1	0	0	0

* reads count of RNA-seq data

Supplementary Table S2. has-miR-941 expression pattern in several human tissues and cell lines

Tissues or cell-lines	Reads (counts)	Reads (TPM) *	Platform	Data Source
hESCs	146	62	Illumina	Morin et al., 2008, Genome Research.
hEBs	63	30	Illumina	Morin et al., 2008, Genome Research.
Fibroblast	39	7	Illumina	SRR037876
Liver	39	7	Illumina	GSM531974
Endometrium	137	5	Illumina	GSM651905
prostate	93	7	Illumina	GSM605626
Prefrontal Cortex	31	10	Illumina	GSE26545
Cerebellum	115	28	Illumina	GSE26545
SW480	130	63	Illumina	GSM416758
MB-MDA231	109	57	Illumina	GSM416761
U2OS	102	52	Illumina	GSM416754
DLD2	77	56	Illumina	GSM416759
Hela	126	56	Illumina	GSM416753
A549	54	34	Illumina	GSM416756
143B	172	95	Illumina	GSM416755
MCF7	89	40	Illumina	GSM416760
HEK293	119	97	Illumina	GSM416733
six human tonsillar B cell populations	1138	44	Illumina	GSE23090
H520	280	208	Illumina	GSM416757
THP-1 (AGO1)	61	54	Illumina	DRR000560
THP-1 (AGO2)	98	20	Illumina	DRR000561
THP-1 (AGO3)	55	39	Illumina	DRR000562
Jurkat cells (AGO2)	386	5596	454	Azuma-Mukai et al., 2008, PNAS.
Jurkat cells (AGO3)	95	1108	454	Azuma-Mukai et al., 2008, PNAS.
H1 hESCs (AGO2)	detected	detected	SOLiD	Goff et al., 2009, PLoS One.

* The expression was normalized by total mapped reads as Transcripts Per Million (TPM)

Supplementary Table S3. The PCR information

Species	Tissue	PCR primer first round	PCR primer second round	Sequencing primer
Chimpanzee	Prefrontal Cortex	Forward: 5'GCTGTGATGGCACCGACGTGT3' Reverse: 5'CCCGGTCCGACGCAGGACTA 3'		
Chinese Rhesus macaque	Prefrontal Cortex	Forward: 5'GGCACCGACGTGTGTCCAGG 3' Reverse: 5'TGGTCCGACGCAGGACGACT 3'		
Human	Prefrontal Cortex	Forward: 5' TCCCAGGTTTACACCATTC 3' Reverse: 5' TCCAGCGTATCCCAGTCC 3'	Forward: 5' ACGTGTCCGGGGAGAGGACG 3' Reverse: 5' CCCGGTCCGACGCAGGACTA 3'	Forward: 5' ACGTGTCCGGGGAGAGGACG 3' Reverse: 5' CCCGGTCCGACGCAGGACTA 3'

Supplementary Table S4. Target effect of miRNA transfection experiments using 6 miRNA target prediction tools

Target prediction tools	Target number	Median value of inhibition (HSF2)	Kolmogorov-Smirnov test p-value (HSF2)	Median value of inhibition (293T)	Kolmogorov-Smirnov test p-value (293T)	Median value of inhibition (HEK293)	Kolmogorov-Smirnov test p-value (HEK293)
TargetMiner	144	-0.028	0.020341017	-0.044	6.41E-05	-0.07	1.70E-05
mirDB - MirTarget2	44	-0.098	9.21E-06	-0.081	0.001204202	-0.122	7.24E-05
2step-SVM (score>0)	134	-0.09	2.14E-08	-0.066	6.43E-08	-0.115	8.27E-11
mirSVR	304	-0.061	3.68E-10	-0.004	0.001383571	-0.039	1.15E-05
TargetScan (all)	371	-0.057	1.42E-11	-0.024	2.77E-07	-0.043	1.03E-09
TargetScan (25 percentile) *	277	-0.068	1.96E-11	-0.03	1.55E-08	-0.067	5.71E-11
TargetScan (50 percentile)	188	-0.084	2.04E-11	-0.061	7.42E-10	-0.073	1.15E-08
TargetScan (75 percentile)	99	-0.089	1.96E-06	-0.095	5.08E-08	-0.144	4.47E-09
PITA (ALL, 3/15 flank)	254	-0.037	0.000205559	-0.02	0.001889136	-0.063	2.98E-08

* TargetScan predicted targets were ranked based on context score.

Supplementary Table S5. miR-941 target effect in prefrontal cortex and cerebellum

Prefrontal cortex (PFC)					
FDR Cutoff	**H-S_H ***	**H-S_L**	**C-S_H**	**C-S_L**	**binomial test**
5%	3	6	5	2	0.02
10%	5	8	6	3	0.03
20%	9	13	11	10	0.19
All targets	57	86	50	46	0.004
Cerebellum (CB)					
FDR Cutoff	**H-S_H**	**H-S_L**	**C-S_H**	**C-S_L**	**binomial test**
5%	6	5	8	0	0
10%	10	9	9	2	0.003
20%	17	20	16	8	0.01
All targets	74	104	70	40	2.787E-09

* H-S_H: The genes with human specific high expression pattern.
H-S_L: The genes with human specific low expression pattern.
C-S_H: The genes with chimpanzee specific high expression pattern.
C-S_L: The genes with chimpanzee specific low expression pattern.

Supplementary Table S6. The genes with species-specific miR-941 target sites gain and loss

Human-specific loss	Human-specific gain	Chimpanzee-specific loss	Chimpanzee-specific gain
GLTSCR1	HK3	PTCD1	NHLH1
ABHD8	KRT6B	GINS4	KIAA0562
GIPC1	VAPB	DNAJB2	ZNF572
ICAM4	VCAM1		GBP2
JMJD7	CCDC140		C22orf24
DNAJC16	GALNT4		B4GALT7
KCNN1	RSPH1		RPGRIP1L
SEPT5	BTBD7		PARP16
	LIX1		PRKAB1

Supplementary Table S7. Population information for HGDP-CEPH Human Genome Diversity Cell Line Panel

Index *	Population	Continents
1	Bantu	Africans
2	Yoruba	Africans
3	San	Africans
4	Mozabite	Africans
5	Orcadian	Europeans
6	Basque	Europeans
7	French	Europeans
8	Italian	Europeans
9	Sardinian	Europeans
10	Tuscan	Europeans
11	Bedouin	Western asians
12	Druze	Western asians
13	Palestinian	Western asians
14	Makrani	Central and Southern Asians
15	Sindhi	Central and Southern Asians
16	Pathan	Central and Southern Asians
17	Burusho	Central and Southern Asians
18	Hazara	Central and Southern Asians
19	Kalash	Central and Southern Asians
20	Han	Eastern Asians
21	Lahu	Eastern Asians
22	Tu	Eastern Asians
23	Cambodian	Eastern Asians
24	Japanese	Eastern Asians
25	Yakut	Eastern Asians
26	Melanesian	Oceanians
27	Papuan	Oceanians
28	Karitiana	Native Americans
29	Surui	Native Americans

* The index are corresponding to Figure3b.

Supplementary Table S8. miR-941 expression estimation in human PFC and CB by Northern Blot.

	CB	PFC	Fold (PFC/CB)
RNU6	160.661[a]	121.25	
miR-941	9.048[b]	8.345	1.22[c]
	40.145	35.545	1.17
	57.288	52.819	1.21
	70.761	66.053	1.25[d]

[a] The expression of RNU6 and miR-941 were estimated using Quantity One software from Bio-Rad Laboratories.

[b] miR-941 expression were estimated under four different exposure times.

[c] The relative abundance of miR-941 between PFC and CB are normalized using RNU6 as a reference.

[d] The relative abundance of miR-941 between PFC and CB corresponding to bands shown on Fig. 1d.

Supplementary Methods

miRNA 5' heterogeneity and cytoplasm/nucleus enrichment analysis

miRNA 5' heterogeneity was estimated as described elsewhere[1]. Briefly, all RNA-seq reads mapping within five nucleotides upstream or downstream of the annotated 5'-position of the mature miRNAs were retained. Then, for each mature miRNA, the sequence with a maximal copy number was designated as the reference sequence. Finally, the heterogeneity of its termini was calculated as the mean of the absolute distances between the observed 5'- or 3'- ends and the ends of the reference sequence. Cytoplasm/nucleus enrichment analysis of miR-941 mature sequence was based on the data from THP-1 (Human acute monocytic leukemia cell line) as described in[2].

miRNA-941 Sequence Evolution analysis

Number of miR-941 precursors in the reference human genome was estimated by mapping annotated miR-941 precursor sequences to the genome (UCSC genome accession code hg19) using BLAST or BLAT. RNA secondary structures of the human miR-941 precursor and corresponding regions in the genomes of chimpanzee, Indian and Chinese rhesus macaques and Denisova were analyzed by RNA-fold[3]. Genomic locations of miR-941 mature sequence and miR-941-star sequence were determined by mapping RNA-seq reads to the miR-941 precursor sequence. Number of miR-941 precursors in Denisova was estimated by mapping publicly available Denisova sequence reads to the human reference genome. Specifically, we mapped Denisova genome reads to the human miR-941 reference genome region containing seven potential precursor sequences (chr20: 62550778-62551487). Out of seven, two were covered by at least two sequence reads at every nucleotide position. The locations of these two precursor sequences within the reference human genome are chr20: 62550795-62550884 and chr20: 62551270-62551359. To determine miR-941 precursor copy number in humans and verify its absence in the chimpanzee and rhesus macaque genomes, we amplified and sequenced the miR-941 genomic locus from the genomes of one human, eight chimpanzee and six rhesus macaque individuals. Sample and primer information used in this analysis is listed in Supplementary Table S3.

miR-941 Precursor Copy Number Variation Analysis

To determine miR-941 precursor copy number variation among human populations we amplified and sequenced genomic region containing miR-941 precursor sequences in 558 individuals from 38 populations of the HGDP-CEPH Human Genome Diversity Cell Line Panel[4] (Supplementary Table S7). Briefly, A two round nested PCR was performed using 10ng of genomic DNA for reaction mixture (50 μl) containing 1.25 U/reaction of OneTaq DNA polymerase (NEB) and 0.2 μM of each primer. Two pair of primers (Supplementary Table S3) were used in amplification carried out at 62 °C and 60 °C as annealing temperature respectively. The final PCR products were electrophoresed, purified and sequenced by an ABI 3730. PCR amplification in the remaining individuals from the HGDP-CEPH panel has failed for technical reasons.

miR-941 precursor copy number were estimated by mapping annotated miR-941 precursor sequences to the amplified and sequenced genomic regions using Blat. The miR-941 precursor copy number variation results were robust to use of other copy number quantification procedures: merging overlapped precursors or counting numbers of tandem repeats constituting miR-941 precursor. miR-941 precursor copy number difference among populations from different geographical regions was tested by Kruskal-Wallis test. miR-941 precursor copy number variant difference among populations from different geographical regions and between sub-Saharan Africans and "out of Africa" populations were tested by Bartlett test and Levene test. Mozabite population was classified as "West Asians" rather than "Africans" in the region-based copy number and copy number variation analyses. To confirm the robustness of miR-941 precursor copy number estimates among humans obtained using PCR, we did repeat PCR amplification followed by sequencing for 6 individuals from African populations. In all 6 cases, miR-941 precursor copy number estimates agreed between the experimental replicates (see Supplementary Fig. S8).

miRNA Transfection, Microarray Data Analysis and miR-941 Target Effects

miRNA transfection experiments were conducted in six cell lines - two human derived kidney cell lines (HEK and 293T), one human skin fibroblast cell line (HSF2), two macaque derived kidney cell lines (LLCMK2 and FrhK-4), one macaque skin fibroblast cell line (MSF) - as described previously in[5]. Briefly, cells were plated in 0.5 ml of growth medium 24h prior to transfection

without antibiotics. miR-941 mimics-Lipofectamine 2000 (Invitrogen) complexes were prepared freshly before transfection based on the manufacturer's protocol. Cells were transfected in six-well plates using miRNA mimics-Lipofectamine 2000 with a final oligonucleotide concentration of 10nmol/L. In parallel, negative control transfections with mock oligonucleotides were conducted according to the manufacturer's protocol. After 24h, cells were harvested and total RNA were extracted using Trizol reagent (Invitrogen) and further processed and hybridized to Affymetrix Human Genome U133 Plus 2.0 arrays following the manufacturer's instructions. R *RMA* package was used to quantify gene expression levels. All original microarray data are deposited in the NCBI GEO database GSE35621.

Ago2 Immunoprecipitation (Ago2-IP) experiments after miR-941 overexpression were conducted in 293T cell line. Briefly, all transfections were performed using human 293T cells cultured in 6-well tissue culture plates. Lipofectamine 2000 (Invitrogen) was used for a Synthetic miR-941 or a scrambled oligo transfection, at 30 nmol/l each (final concentration) per $1x10^6$ cells/well of a 6-well plate using DharmaFECT (GE Healthcare). Total $5x10^6$ cells were collected and subjected to Ago2 immunoprecipitation (Ago2-IP) using the RNA isolation kit Mouse Ago2 (Wako Chemicals) according to the manufacturer's instructions. For a negative control, immunoprecipitation was performed using non-immune IgG beads prepared with the antibody immobilization bead kit (Wako Chemicals). The IP pull down RNA was used as template for an "in vitro" transcription reaction generating biotin-labeled antisense cRNA. The cRNA was analyzed on Affymetrix Human Genome U133 Plus 2.0 arrays following the manufacturer's instructions. The R *RMA* package was used to quantify gene expression levels. All original microarray data are deposited in the NCBI GEO database GSE35621.

We used GOstats[6] to investigate putative functions of experimentally verified target genes of miR-941 in human cell lines based on our transfection results. Experimentally verified target genes using miR-941 transfection experiments were predicted by combining six target prediction algorithms plus genes that with at least two miR-941 seed region matches (2-7nt from 5'terminus) within 3'UTR and were required to show down-regulation by miR-941 transfection on the average of three cell lines (Supplementary Data 2). Experimentally verified target genes were used in functional enrichment analyses, the other genes expressed in the three cell lines were used as a background. Significantly enriched KEGG pathways and target genes in these pathways were shown in Figure4 (Hypergeometric test, Bonferroni corrected $p<0.05$). The putative direct target

genes of miR-941 were obtained using Ago2 Immunoprecipitation (Ago2-IP) experiments. The direct target genes were predicted by combining six target prediction algorithms, as well as genes containing at least two miR-941 seed region matches (2-7nt from 5'terminus) within their 3'UTRs, and were required to show at least 0.15 log2 fold changes compared with negative control (Supplementary Data 2).

Evolution of miRNA-941-guided Regulation

The approach to calculating miR-941 regulation effect on human-specific gene expression changing in prefrontal cortex and cerebellum was adopted and modified from[5]. Briefly, experimentally verified target genes were identified based on miR-941 transfection experiments. Inhibition ratio was calculated as the ratio between the number of experimentally verified miR-941 target genes with lineage specific low expression to that with lineage specific high expression on human and chimpanzee in prefrontal cortex and cerebellum. Human inhibition excess is tested by comparing to the inhibition ratio in chimpanzee using Binomial test. We also tested miR-941 regulation effect based on experimentally verified target genes screened based on different FDR cutoffs more stringently. Specifically, we calculated proportions of predicted target genes and non-target genes inhibited after transfection in both cell lines, at a certain inhibition cutoff (calculated as the difference in expression between miRNA transfection and the negative control). FDR was calculated as the ratio of the proportion of non-target genes passing this inhibition cutoff compared to the total proportion of target genes expressed in the corresponding cell line. The unions of targets were used as experimental verified miRNA targets at FDR 5%, 10%, 20% cutoffs (Supplementary Data 2). Binomial test were used to test whether the human shows inhibition excess in prefrontal cortex and cerebellum (Supplementary Table S5).

Species-specific gain/loss of miR-941 target sites was estimated using binding site predictions by TargetScan based on human, chimpanzee and rhesus macaque 3'UTR sequence alignments. Specifically, 3'UTR sequence alignments of human, chimpanzee and rhesus macaque were extracted from 3'UTR sequence alignment file downloaded from TargetScan website[7]. Only the alignments with less than 5% of the gap sequence in human, chimpanzee and macaque were used for downstream analysis. TargetScan was used to predict miR-941 target sites across three species. Gain/loss of the target sites on the human and the chimpanzee lineages was calculated using rhesus macaque sequence as an outgroup. Human-specific gain (HSG) ratio of miR-941 target

genes was calculated as the ratio between human-specific gained miR-941 target gene number and the total number of target genes gained on the human and the chimpanzee lineages. Human-specific loss (HSL) ratio was calculated as the ratio between the number of human-specific loss of miR-941 target genes and the total number of target genes lost on the human and the chimpanzee lineages.

Supplementary References

59. Seitz, H., Ghildiyal, M. & Zamore, P.D. Argonaute loading improves the 5' precision of both MicroRNAs and their miRNA* strands in flies. *Curr Biol* **18**, 147-151 (2008).
60. Liao, J.Y. et al. Deep sequencing of human nuclear and cytoplasmic small RNAs reveals an unexpectedly complex subcellular distribution of miRNAs and tRNA 3' trailers. *PLoS One* **5**, e10563.
61. McCaskill, J.S. The equilibrium partition function and base pair binding probabilities for RNA secondary structure. *Biopolymers* **29**, 1105-1119 (1990).
62. Cavalli-Sforza, L.L. The Human Genome Diversity Project: past, present and future. *Nat Rev Genet* **6**, 333-340 (2005).
63. Hu, H.Y. et al. MicroRNA expression and regulation in human, chimpanzee, and macaque brains. *PLoS Genet* **7**, e1002327.
64. Beissbarth, T. & Speed, T.P. GOstat: find statistically overrepresented Gene Ontologies within a group of genes. *Bioinformatics* **20**, 1464-1465 (2004).
65. Grimson, A. et al. MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol Cell* **27**, 91-105 (2007).